Question **1**
Not yet answered
Marked out of 1.00

Which of the following is **correct** regarding crowdsourcing?

Select one:

- a. Honey pots can detect uniform spammers, random spammers and sloppy workers.
- b. Majority Decision and Expectation Maximization both give less weight to spammers' answers.
- c. The accuracy of majority voting is never equal to the one of Expectation Maximization.
- d. Uniform spammers randomly select answers.

Question **2**
Not yet answered
Marked out of 1.00

Which of the following is **correct** regarding prediction models?

Select one:

- a. Simple models have higher bias than complex models.
- b.
  A high bias is a sign of overfitting.
- c. In low data regime, complex models tend to perform better.
- d. A high variance is a sign of underfitting.

Question **3**
Not yet answered
Marked out of 1.00

In classification, which of the following is **true** regarding class imbalance?

Select one:

- a. Oversampling the larger class can reduce the impact of the skewed distribution.
- b. The leave-one-out methodology produces the same class distribution in the training and the testing set.
- c. Oversampling rare classes in the testing set can reduce the impact of skewed distribution.
- d. Classes should have the same distribution in the validation set and in the full dataset.

Question **4**
Not yet answered
Marked out of 1.00

Which of the following is **correct** regarding the use of *Hidden Markov Models (HMMs)* for entity recognition in text documents?

Select one:

- a. The cost of predicting a word is linear in the lengths of the text preceding the word.
- b. The label of one word is predicted based on all the previous labels
- c. The cost of learning the model is quadratic in the lengths of the text.
- d. An HMM model can be built using words enhanced with morphological features as input.

Question **5**

Not yet answered

Marked out of 1.00

Which of the following is **true** for community detection in social graphs?

Select one:

- a. The result of the Girvan-Newman algorithm can depend on the order of processing of nodes whereas for the Louvain algorithm this is not the case.
- b. The Louvain algorithm is efficient for small networks, while the Girvan-Newman algorithm is efficient for large networks.
- c. If n cliques of the same order are connected cyclically with n-1 edges, then the Louvain algorithm will always detect the same communities, independently of the order of processing of the nodes.
- d. The Louvain algorithm always creates a hierarchy of communities with a common root.

Question **6**

Not yet answered

Marked out of 1.00

Which is a **correct** pruning strategy for decision tree induction?

Select one:

- a. Build the full tree, then replace subtrees with leaf nodes labelled with the majority class, if classification accuracy does not change.
- b. Remove attributes with lowest information gain.
- c. Stop partitioning a node when the number of positive and negative samples are equal.
- d. Apply Maximum Description Length principle.

Question **7**

Not yet answered

Marked out of 1.00

If for the $\chi 2$ statistics for a binary feature we obtain $P(\chi 2 \mid DF = 1) < 0.05$ this means

Select one:

- a. That the class label correlates with the feature
- b. That the two features are correlated.
- c. That the class label is independent of the feature
- d. None of the above

Question **8**

Not yet answered

Marked out of 1.00

Which of the following is **true** for Recommender Systems (RS)?

Select one:

- a. Matrix Factorization can predict a score for any user-item combination in the dataset.
- b. Item-based RS need not only the ratings but also the item features
- c. The complexity of the Content-based RS depends on the number of users
- d. Matrix Factorization is typically robust to the cold-start problem.

Question **9**

Not yet answered

Marked out of 1.00

In User-Based Collaborative Filtering, which of the following is **correct**?

Select one:

- a. Pearson Correlation Coefficient and Cosine Similarity have different value ranges, but return the same similarity ranking for the users
- b. Pearson Correlation Coefficient and Cosine Similarity have the same value range and return the same similarity ranking for the users.
- c. Pearson Correlation Coefficient and Cosine Similarity have different value ranges and can return different similarity ranking for the users
- d. Pearson Correlation Coefficient and Cosine Similarity have the same value range, but can return different similarity ranking for the users

Question **10**

Not yet answered

Marked out of 1.00

Which of the following is **true** regarding the random forest classification algorithm?

Select one:

- a. We compute a prediction by randomly selecting the decision of one weak learner.
- b. It produces a human interpretable model.
- c. It uses only a subset of features for learning in each weak learner.
- d. It is not suitable for parallelization.

Question **11**

Not yet answered

Marked out of 1.00

Information extraction:

Select one:

- a. Is used to identify characteristic entities in a document.
- b. Necessarily requires training data.
- c. Is always bootstrapped by using ontologies.
- d. Can be used to populate ontologies.

Question **12**

Not yet answered

Marked out of 1.00

Which of the following is **false** for Ontologies?

Select one:

- a. They give the possibility to specify schemas for different domains.
- b. They dictate how semi-structured data are serialized.
- c. Different information systems need to agree on the same ontology in order to communicate.
- d. They help in the integration of data expressed in different models.

Question **13**

Not yet answered

Marked out of 1.00

Which of the following properties is part of the RDF Schema Language?

Select one:

- a. Predicate
- b. Type
- c. Domain
- d. Description

◀ Quiz 1

Jump to...

Contact EPFL CH-1015 Lausanne +41 21 693 11 11

Follow EPFL on social media